

DOCUMENT RESUME

ED 073 926

SE 015 809

AUTHOR Beardslee, Edward C.; Jerman, Max E.
TITLE Linguistic Variables in Verbal Arithmetic Problems.
PUB DATE [73]
NOTE 26p.
EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS *Elementary School Mathematics; Linguistics;
*Mathematics Education; *Problem Solving; Reading;
*Research; Secondary School Mathematics; Structural Analysis

ABSTRACT

This paper describes a study in which 14 linguistic variables were used to determine which variables would account for a significant amount of the observed variance in the error rate in verbal arithmetic problems. Three forms of verbal problem sets in which the number of words in the problem statement were systematically varied were administered to classes of students in grades four through eight. Regression analysis showed that none of the variables accounted for a significant amount of variance for all grades, although four variables did enter the regression within the first six steps on two or more of the test forms for most grades. Regression analysis on a selected subset of six variables produced results similar to those provided by an analysis involving all 14 original variables. (Author/DT)

Abstract

Linguistic Variables in Verbal
Arithmetic Problems

by

Edward C. Beardslee
and
Max E. Jerman

The Pennsylvania State University

Fourteen linguistic variables, defined in a previous study, were used to determine which variables would account for a significant amount of the observed variance in the error rate in verbal arithmetic problems. In this study, three forms of a verbal problem set in which the number of the words in the problem statements were systematically varied were administered to classes of students in Grades 4 - 8. Using regression analysis, none of the variables were found to account for a significant amount of the variance for all grades, although four variables did enter the regression within the first six steps on two or more of the test forms for most grades. Selecting a subset of six variables, regression analysis produced results which were similar to those provided by the analysis involving all fourteen original variables.

ED 073926

5-15 809
ERIC
Full Text Provided by ERIC

Linguistic Variables in Verbal
Arithmetic Problems

By

Edward C. Beardslee
and
Max E. Jerman
The Pennsylvania State University

Jerman and Rees (1972) and Jerman (1973a, 1973b) investigated the influence of structural variables on the relative difficulty of verbal arithmetic problems for students in grades 4 - 9 (1972, 1973b) and in grades 4 - 8 (1973a). Krushinski (1973) investigated the influence of certain linguistic variables on the relative difficulty of verbal arithmetic problems for college students enrolled in a methods of teaching elementary mathematics course. One purpose of this study was to determine the influence of the linguistic variables on the ability of students to solve verbal arithmetic problems in grades 4 - 8. A second purpose was to identify a small subset of the variables (no more than six) which could be used in further investigations.

The data from Jerman (1973a) were used in this investigation. In that study, Jerman determined that several structural variables which had accounted for a significant amount of the variance in the linear regression in his

previous study (1973b) also accounted for a significant amount of the variance in the more recent study (1973a). In addition, the number of words in the problem statement entered the regression in a few of the analyses but not all of them. Hence, Jerman suggested that it is not simply the number of words in the problem statement that influence its difficulty, but the number of words in relation to other factors. Thus an attempt to determine these "other factors" was undertaken. Krushinski (1973) working with a set of 14 linguistic variables found that the following six; sentence length, the number of clauses, clause length, the number of prepositional phrases, the number of words in the question sentence and the numeral in the question sentence accounted for a significant amount of the variance in his study with college students. The fourteen linguistic variables used in the regression analysis in the present study were the same as those used by Krushinski and were defined as follows:

1. Length (LENGTH) was defined as the number of words in the problem statement. A count of one was assigned to a single word, hyphenated word, or group of words that would appear as a single entry in a dictionary. For example, P.M. was counted as 1. Also, numerals were converted to words and then counted. For example, \$3.75 was assigned a count of 5 (\$3.75 = three dollars and seventy-five cents).

2. Sentences (SENT) was defined as the number of sentences in the problem.
3. Sentence length (SENTLN) was defined as the average number of words per sentence and was obtained by forming the ratio of variables 1 and 2 respectively:
$$\text{SENTLN} = \frac{\text{LENGTH}}{\text{SENT}} .$$
4. Main clauses (MAINCL) was defined as the number of main clauses in the problem.
5. Subordinate clauses (SUBCL) was defined as the number of subordinate clauses in the problem.
6. Clauses (CLAUSE) was defined as the total number of clauses in the problem and was obtained by adding the values of variables 4 and 5:
$$\text{CLAUSE} = \text{MAINCL} + \text{SUBCL} .$$
7. Words in the main clauses (WDMAIN) was defined as the number of words in the main clauses (identified by MAINCL) in the problem.
8. Words in the subordinate clauses (WDSUB) was defined as the number of words in the subordinate clauses (identified by SUBCL) in the problem.
9. Clause length (CLSLN) was defined as the average number of words per clause and was obtained by forming the ratio of variables 1 and 6 respectively:
$$\text{CLSLN} = \frac{\text{LENGTH}}{\text{CLAUSE}} .$$

10. Main clause length (MCLSLN) was defined as the average number of words per main clause and was obtained by forming the ratio of variables 7 and 4 respectively:

$$MCLSLN = \frac{WDMAN}{MAINCL} .$$

11. Subordinate clause length (SCLSLN) was defined as the average number of words per subordinate clause and was obtained by forming the ratio of variables 8 and 5 respectively:

$$SCLSLN = \frac{WDSUB}{SUBCL} .$$

12. Number of prepositional phrases (PREPHR) was defined as the number of prepositional phrases in the problem.
13. Number of words in the question sentence (WDSHQ); one count was given for each word in the question sentence in the problem.
14. Numeral in question sentence (NUMINQ) was given a value of 1 if the question sentence contained numerals and a value of 0 otherwise.

Method

Three test forms of 30 problems each were prepared using the problem set of 30 problems employed in the Jerman (1973b) study. Three types of each problem were prepared. Type 1 was the original problem set with one-third

fewer words in each problem. Type 2 was the original problem set. Type 3 contained one-third more words in each problem than the original problem set. The digits in each problem, the order of operations, the cues, and all other structural aspects (as pertaining to the structural variables defined by Jerman, 1973a) were held constant, however extra clauses and modifiers were added to Type 2 problems to qualify them as Type 3 problems. Type 1 problems were obtained by eliminating words and phrases from Type 2 problems. Each of the three test forms contained 30 problems; 10 of Type 1, 10 of Type 2, and 10 of Type 3. While the problems were distributed among the three forms on a random basis, they were arranged in the same order on each test, from easy to more difficult on the basis of previous experience.

In May 1971, the problem sets were administered to students by their classroom teachers. The three test forms were randomized prior to distributing them to the students. Students performed all their work on the test forms and were completed during their normal mathematics period which ranged from 40 to 50 minutes in length. All of the tests were scored by Jerman.

Results

The mean percentage correct and standard deviation for each class group are presented in Table 1. For a

detailed discussion concerning the implications of these figures and a statistical comparison of the means, the reader is referred to Jerman (1973a). For purposes of this study, the focus is on the linguistic variables in a regression model.

Insert Table 1 about here

Tables 2, 3, and 4 present the number of each variable which entered at each step in the regression for each group for each of the three test forms. The multiple R for the sixth step and the total R are also given.

Insert Tables 2, 3, 4 about here

Observing the tables, the two variables which entered within the first several steps with the greatest frequency were variables 8 (WDSUB) and 12 (PREPHR). In fact, the number of words in the subordinate clauses (WDSUB) entered first in 9 of the 18 analyses. Also, the number of prepositional phrases (PREPHR) entered second in 9 of the 18 regressions. In particular, observing the totals columns for each of the three forms reveals that variable 8, (WDSUB) entered first, followed by variable 12 (PREPHR) for forms 2 and 3 but neither variable entered the regression within the first six steps for Form 1.

The regression coefficients, standard errors of regression coefficients, and computed t-values for each of the variables in the first six steps for each group are shown in Tables 5, 6 and 7.

Insert Tables 5, 6, 7 about here

Variable 8 (WDSUB) was significant for several of the groups who received Form 3 of the test but for only 1 group in each of Forms 1 and 2. Four other variables, 3 (SENTLN), 5 (SUBCL), 9 (CLSLN) and 10 (MCLSLN) were significant for two of the 18 analyses while 12 (PREPHR) was significant for only 1 of the groups. Referring to the "Total" column of Tables 5, 6, and 7 which shows the first six variables which entered the regression based on the mean percentage correct for all students, it is worth noting that only one variable, 3 (SENTLN), entered the regression for all three forms of the test.

Based on the regression analyses discussed above, an attempt was made to reduce the number of variables without sacrificing much of the variance accounted for by the fourteen variables. With a limit of six variables in mind, several analyses were performed which resulted in six variables being selected. These were: 3 (SENTLN), 5 (SUBCL), 8 (WDSUB), 9 (CLSLN), 12 (PREPHR), and 14 (NUMINQ).

Tables 8, 9, and 10 present the number of each of the six variables which entered at each step in the regression for each group for each of the three test forms. The multiple R for the fourth step and the total R are also given.

Insert Tables 8, 9, 10 about here

Comparing the R's in Tables 8, 9, and 10 with those in Tables 2, 3, and 4, one observes that the difference between the R at the Sixth step on Tables 2, 3, and 4 and the total R on Tables 8, 9, and 10 exceeds .05 in only one instance (Grade 6, Form 1 where the difference is .104). The greatest discrepancies occur for Form 1 while the fit for Forms 2 and 3 are quite similar.

Tables 11, 12, and 13 present the information for the six variables which corresponds to Tables 5, 6, and 7 for the fourteen variables with the exception that the coefficients that appear in Tables 11, 12, and 13 are for the final regression equation, not the equation after Step 6 as is the case in Tables 5, 6, and 7.

Insert Tables 11, 12, 13 about here

With only the selected six variables in the regression analysis, four variables were significant for one or more of the groups. They were; 3 (SENTLN), 5 (SUBCL), 8 (WDSUB)

and 9 (CLSLN). Observing the "Total" column in Tables 11, 12, and 13, note that for Form 1, variables 3 and 12 failed to enter the regression at all while they entered first and second respectively for both of the other test forms.

Discussion

Only two of the six variables which Krushinski (1973) found to be significant entered consistently among the first six in the linear regression in this study; these were, 9 (CLSLN) and 12 (PREPHR). CLSLN was significant in only 3 of the 18 separate analyses while PREPHR was significant only once. Since Krushinski's study involved college students and the students in the present study were from grades 4 - 8, the fact that two variables appear to be important in the regression analysis with both groups raises hopes that it is possible to find specific linguistic variables which can be shown to contribute to the difficulty of solving word problems for all students.

Encouraging also, is that four variables: 3 (SENTLN), 8 (WDSUB), 9 (CLSLN), and 12 (PREPHR); entered the regression consistently within the first six steps on two or more of the test forms. This result suggests that specific linguistic variables can be identified which will contribute to the difficulty level of verbal arithmetic problems on all test forms. On the other hand, several variables used in this

study entered the regression consistently on only one of the three test forms; these were, variables 5 (SUBCL), 10 (MCLSLN), and 14 (NUMINQ) on Form 1 and variable 11 (SCLSLN) on Form 2. The failure of these variables to contribute to the regression on more than one test form suggests that variables other than those identified in this study influenced the difficulty level of the problems on the three test forms.

Noticeably absent from being significant in the regression was the length variable (LENGTH). Earlier studies, (Jerman 1973a, 1973b), found that the variable for the number of words in the problem statement was significant. In the present study, the number of words was not significant. In fact, LENGTH entered among the first six variables on only five occasions--coming in no higher than step 3. Perhaps one of the explanations for the discrepancy in the results is the fact that the definitions used were different. The definition used in this study counted the word equivalent of numerals while the previous studies assigned a count of 1 to each numeral. This subtle difference coupled with the fact that NUMINQ (numeral in the question sentence) appears to be an important variable suggests that more attention be paid to any numerals in the problem statement.

Any broad generalizations based on the results of this study should be made with extreme caution for the simple reason that the multiple R at the end of six steps for the 14 variables and the total R for six variables was .778 or less which indicates that the per cent of

variance accounted for by the variables included in the regression (R^2) was 60% or less. Hence, the linguistic variables identified in the present study do not account for as large a portion of the variance as would be necessary in order to attempt to accurately predict the difficulty level of verbal problems.

References

German, M. Problem Length as a Structural Variable in Verbal Arithmetic Problems. Educational Studies in Mathematics, 1973a (In press).

German, M. Structural Variables in Verbal Arithmetic Problems. Journal for Research in Mathematics Education, 1973b (In press).

German, M. and Rees, R. Predicting the Relative Difficulty of Verbal Arithmetic Problems. Educational Studies in Mathematics, 1972, 4, 306-323.

Krushinski, J. An Analysis of Linguistic Structural Variables that Contribute to Problem Solving Difficulty. Unpublished master's paper, The Pennsylvania State University, 1973.

TABLE 1
Mean and Standard Deviation for Percentage Correct

Test Form									
Grade	1			2			3		
	\bar{X}	S	N	\bar{X}	S	N	\bar{X}	S	N
4	35.71	23.67	28	35.17	21.19	38	37.83	23.42	31
5	42.82	22.47	32	44.97	22.91	35	45.56	23.53	27
6	70.36	22.10	31	59.41	21.44	35	66.10	19.30	40
7	59.19	20.16	75	60.20	21.52	85	61.16	19.40	75
8	69.40	19.17	63	66.20	21.02	60	73.00	18.22	65
Total	58.98	19.34	229	56.84	20.99	253	61.42	18.14	238

TABLE 2
Variables Entered at Each Step
for Form 1 of the Test

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1	10	5	10	5	8	5
2	5	9	5	10	10	10
3	14	3	12	12	3	14
4	11	14	9	3	14	3
5	13	13	11	6	12	7
6	9	7	8	14	9	1
7	8	10	13	11	11	9
8	3	11	7	9	5	11
9	5-	1	2	8	13	13
10	6	2	14	2	7	2
11		12	3	7	2	4
12		5-	6	13		12
13		4				
STEP 6R	.458	.743	.553	.630	.499	.601
TOTAL R	.475	.797	.675	.662	.557	.643

TABLE 3
Variables Entered at Each Step
for Form 2 of the Test

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1	8	12	5	8	8	8
2	12	11	12	12	12	12
3	11	3	14	3	2	2
4	4	4	8	11	11	11
5	14	1	3	9	3	3
6	3	7	5-	13	9	14
7	9	5	10	5	4	10
8	2	14	2	7	14	9
9	6	10	4	14	10	4
10	10	2	7	10	13	7
11		4-	5	2		
12		9	11	6		
13		13	13			
STEP 6R	.616	.573	.592	.587	.652	.599
TOTAL R	.637	.600	.633	.610	.682	.629

TABLE 4
Variables Entered at Each Step
for Form 3 of the Test

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1.	3	8	12	8	8	8
2	5	12	8	12	12	12
3	9	13	4	1	13	1
4	1	9	2	13	11	13
5	6	3	14	11	9	9
6	11	10	9	5	3	3
7	2	14	3	9	1	11
8	5-	11	11	3	6	5
9	12	5	6	14	2	14
10	7	4	10	10	10	10
11	13	7	7	8-	4	8-
12	4	2	13	6		2
13	14			13-		1-
14	10			7		8
15				2		7
16						6
STEP 6R	.557	.778	.664	.728	.699	.718
TOTAL R	.631	.795	.808	.805	.730	.773

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 1

[illegible]

TABLE 6

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 2

Variable	Grade 4			Grade 5			Grade 6		
	Coef.	S.E.	t	Coef.	S.E.	t	Coef.	S.E.	t
1	-.009	.036	.250	.021	.028	.750			
2	.049	.130	.377	-.055	.049	1.122	-.016	.019	.842
3				-.180	.212	.849			
4									
5									
6				-.012	.027	.444	.036	.014	2.571*
7	.011	.026	.423						
8									
9									
10	.054	.045	1.200	.049	.050	.980	.056	.035	1.600
11	.076	.054	1.407	.117	.065	1.800			
12									
13	-.109	.248	.440				-.205	.180	1.139
14									
	c = -.304			c = .285			c = -.401		
Variable	Grade 7			Grade 8			Total		
	Coef.	S.E.	t	Coef.	S.E.	t	Coef.	S.E.	t
1	-.056	.047	1.191	.044	.126	.349	.018	.111	.162
2				-.053	.059	.898	-.022	.029	.759
3									
4									
5									
6									
7									
8	.036	.032	1.125	.039	.030	1.30	.019	.020	.950
9	.061	.076	.803	.058	.080	.725			
10									
11	.043	.040	1.075	.024	.038	.632	.032	.035	.914
12	.057	.056	1.018	.068	.049	1.388	.069	.043	1.605
13	-.012	.024	.500						
14							-.119	.193	.617
*p < .05	c = -.661			c = -1.032			c = -.416		

TABLE 7

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 3

Variable	Grade 4			Grade 5			Grade 6		
	Coef.	S.E.	t	Coef.	S.E.	t	Coef.	S.E.	t
1	-.024	.025	.960						
2									
3	-.044	.071	.620	-.067	.047	1.426	.432	.322	1.342
4									
5	.300	.267	1.124				-.457	.314	1.455
6	.165	.245	.673						
7									
8	.168	.098	1.714	.053	.017	3.118**	.020	.008	2.500*
9				.177	.086	2.058	.018	.033	.545
10				-.068	.058	1.172			
11	.026	.041	.634						
12				.053	.072	.736	.070	.064	1.094
13				-.018	.015	1.200			
14							-.139	.165	.842
				c = -1.077			c = -.830		
Variable	Grade 7			Grade 8			Total		
	Coef.	S.E.	t	Coef.	S.E.	t	Coef.	S.E.	t
1	-.010	.009	1.111						
2							-.009	.007	1.286
3									
4	.121	.203	.596	-.106	.047	2.255*	-.053	.037	1.432
5									
6	.013	.025	.520						
7									
8									
9	.023	.032	.719	.043	.016	2.688*	.044	.015	2.933**
10	.122	.051	2.392*	.152	.064	2.375*	.086	.051	1.686
11				.058	.035	1.657			
12				.012	.068	.176	.049	.055	.891
13	-.017	.016	1.063	-.021	.016	1.312	-.001	.013	.077
14									
*p < .05	c = -.451			c = -1.122			c = -.602		
**p < .01									

TABLE 8
 Variables Entered at Each Step
 for Form 1 of the Test,
 6 Selected Variables

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1	9	5	9	5	8	5
2	5	9	5	9	9	9
3	14	3	8	3	14	3
4	8	14	12	14	3	14
5	3	8	3	8	5	
6	12	12		12		
STEP 4R	.400	.738	.443	.578	.448	.566
TOTAL R	.418	.740	.449	.581	.465	.566

TABLE 9
Variables Entered at Each Step
for Form 2 of the Test,
6 Selected Variables

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1	8	12	5	8	8	3
2	12	8	12	12	12	12
3	5	5	14	3	3	14
4	14	3	8	9	9	3
5	9	9	3	14	14	9
6	3		5-		5	
7			9			
STEP 4R	.586	.546	.580	.551	.644	.578
TOTAL R	.590	.551	.592	.553	.654	.589

TABLE 10
 Variables Entered at Each Step
 for Form 3 of the Test;
 6 Selected Variables

STEP	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	TOTAL
1	3	8	12	8	8	8
2	5	12	8	12	12	12
3	9	14	9	14	5	14
4	12	5	3	5	3	9
5	3-	3	14	3	9	3
6	14	9		9	14	5
STEP 4R	.525	.715	.669	.686	.575	.649
TOTAL R	.526	.772	.673	.705	.661	.690

TABLE 11

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 1

Variable	Grade 4 Coef.	Grade 4 S.E.	t	Coef.	Grade 5 S.E.	t	Coef.	Grade 6 S.E.	t
3	-.033	.055	.600	-.081	.037	2.189*	-.020	.056	.357
5	.354	.370	.957	.326	.250	1.304	.504	.399	1.263
8	-.031	.044	.704	.011	.030	.367	.040	.048	.833
9	.078	.064	1.219	.127	.044	2.886**	.125	.070	1.786
12	.023	.061	.377	.011	.041	.268	.033	.065	.508
14	.359	.294	1.221	.226	.199	1.136			
	c = -.285			c = -.510			c = -1.645		
Variable	Grade 7 Coef.	Grade 7 S.E.	t	Coef.	Grade 8 S.E.	t	Coef.	Total S.E.	t
3	-.038	.034	1.118	-.038	.040	.950	-.042	.030	1.400
5	.311	.234	1.329	.175	.259	.676	.211	.076	2.776*
8	-.009	.028	.321	.004	.032	.125			
9	.084	.041	2.049	.074	.044	1.682	.080	.033	2.424*
12	-.011	.038	.289						
14	.204	.186	1.097	.218	.216	1.009	.211	.157	1.344
	c = -.792			c = -.933			c = -.699		

*p < .05

**p < .01

TABLE 12

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 2

Variable	Grade 4 Coef.	Grade 4 S.E.	t	Grade 5 Coef.	Grade 5 S.E.	t	Grade 6 Coef.	Grade 6 S.E.	t
3	-.014	.050	.280	-.037	.052	.712	-.020	.040	.500
5	-.095	.266	.357	-.197	.278	.709			
8	.052	.037	1.405	.058	.038	1.526	.038	.021	1.810
9	.032	.088	.364	.039	.092	.424	.010	.067	.149
12	.042	.053	.792	.090	.054	1.667	.054	.039	1.385
14	-.116	.228	.509				-.204	.184	1.109
	c = -.219			c = -.356			c = -.435		
Variable	Grade 7 Coef.	Grade 7 S.E.	t	Grade 8 Coef.	Grade 8 S.E.	t	Total Coef.	Total S.E.	t
3	-.035	.043	.814	-.055	.042	1.310	-.031	.037	.838
5				-.034	.223	.152			
8	.044	.023	1.913	.061	.031	1.968	.040	.020	2.000
9	.050	.074	.676	.070	.073	.959	.042	.063	.657
12	.036	.043	.837	.049	.045	1.089	.039	.036	1.083
14	.054	.200	.270	-.132	.191	.691	-.112	.171	.655
	c = -.716			c = -.874			c = -.540		

TABLE 13

Regression Coefficients, Standard Errors of Regression Coefficients
and Computed t-values for Form 3

Variable	Grade 4 Coef.	Grade 4 S.E.	t	Grade 5 Coef.	Grade 5 S.E.	t	Grade 6 Coef.	Grade 6 S.E.	t
3									
5	.222	.098	2.265*	-.093	.043	2.163*	-.062	.038	1.632
8				-.242	.204	1.186			
9	.055	.038	1.447	.089	.028	3.178**	.041	.015	2.733*
12	.036	.077	.468	.129	.067	1.925	.117	.059	1.983
14	-.035	.221	.158	.052	.072	.722	.019	.062	.306
				-.230	.176	1.307	-.080	.156	.513
			c = -.654			c = -.561			c = -1.139
Variable	Grade 7 Coef.	Grade 7 S.E.	t	Grade 8 Coef.	Grade 8 S.E.	t	Total Coef.	Total S.E.	t
3	-.037	.034	1.088	-.090	.042	2.143*	-.048	.032	1.500
5	-.098	.162	.605	-.157	.196	.801	-.084	.151	.556
8	.046	.023	2.000	.068	.027	2.518*	.044	.021	2.095*
9	.050	.053	.943	.130	.064	2.031	.075	.049	1.531
12	.047	.057	.824	-.009	.069	.130	.029	.053	.547
14	-.149	.140	1.064	.030	.170	.176	-.080	.131	.611
			c = -.660			c = -1.062			c = -.734